CLAIMS

We claim:

1. An automatic method for generating summaries for text documents, comprising steps of:

generating a set of sentences for a set of documents by document discourse analysis and a set of words by morphologic process;

initializing a score for each word in the set of words and for each sentence in the set of sentences;

computing the score for each word in the set of words according to the score of sentences containing it and the correlation degree between the word and the user information;

computing the score for each sentence in the set of sentences according to the score of words composing it and the position of the sentence in a section and a paragraph; and

if the sum of scores of the words and the sum of scores of the sentences change apparently, go back to the step of computing the word score; otherwise continuing;

outputting the top-ranked sentences as the summary of the set of documents, the top-ranked words as the keywords list of the set of documents.

2. An automatic method according to claim 1, wherein the step of computing the score for each word comprises:

computing the score for each word in the set of words according to the linguistic salience of the word to the user profile.

3. An automatic method according to claim 1, wherein the step of computing the score for each word comprises:

computing the score for each word in the set of words according to the similarity among the word, the query and topic provided by a user.

4. An automatic method according to claim 1, wherein the step of computing the score for each word comprises:

computing the score for each word in the set of words according to the similarity among the word and the terms in the titles of the documents.

5. An automatic method according to claim 1, wherein the step of computing the score for each word comprises:

computing the score for each word in the set of words according to the ratio of its occurrence number in the document to its occurrence number in the set of documents.

6. An automatic method according to claim 1, wherein the step of computing the score for each word comprises:

computing the score for each word in the set of words according to the ratio of the number of the documents including the word to the number of documents in the set of documents.

7. An automatic method according to claim 1, wherein the step of computing the score for each word comprises:

computing the score for each word in the set of words according to the weighted-average of at least two of:

the linguistic salience of the word to the user profile;

the similarity among the word, the query and topic provided by a user;

the similarity among the word and the terms in the titles of the documents;

the ratio of its occurrence number in the document to its occurrence number in the set of documents; and

the ratio of the number of the documents comprising the word to the number of documents in the set of documents.

8. A computer program product for automatically generating summaries for text documents, said computer program product comprising a computer usable medium having computer readable program code thereon, said computer readable program code comprising:

computer program code means for generating a set of sentences for a set of documents by document discourse analysis and a set of words by morphologic process;

computer program code means for initializing a score for each word in the set of words, and each sentence in the set of sentences;

computer program code means for computing the score for each word in the set of words according to the score of sentences containing it and the correlation degree between the word and the user information;

computer program code means for computing the score for each sentence in the set of sentences according to the score of words composing it and the position of the sentence in a section and a paragraph;

computer program code means for determining if the sum of scores of the words and the sum of scores of the sentences exhibit an apparent change; and

computer program code means for outputting the top-ranked sentences as the summary of the set of documents, the top-ranked words as the keywords list of the set of documents.

9. A computer program product for automatically generating summaries according to claim 8, wherein the computer program code means for computing the score for each word comprises:

computer program code means for computing the score for each word in the set of words according to the linguistic salience of the word to the user profile.

10. A computer program product for automatically generating summaries according to claim 8, wherein the computer program code means for computing the score for each word comprises:

computer program code means for computing the score for each word in the set of words according to the similarity among the word, the query and topic provided by a user.

11. A computer program product for automatically generating summaries according to claim 8, wherein the computer program code means for computing the score for each word comprises:

computer program code means for computing the score for each word in the set of words according to the similarity among the word and the terms in the titles of the documents.

12. A computer program product for automatically generating summaries according to claim 8, wherein the computer program code means for computing the score for each word comprises:

computer program code means for computing the score for each word in the set of words according to the ratio of its occurrence number in the document to its occurrence number in the set of documents.

13. A computer program product for automatically generating summaries according to claim 8, wherein the computer program code means for computing the score for each word comprises:

computer program code means for computing the score for each word in the set of words according to the ratio of the number of the documents comprising the word to the number of documents in the set of documents.

14. A computer program product for automatically generating summaries according to claim 8, wherein the computer program code means for computing the score for each word comprises:

computer program code means for computing the score for each word in the set of words according to the weighted-average of at least two of:

the linguistic salience of the word to the user profile,

the similarity among the word, the query and topic provided by a user,

the similarity among the word and the terms in the titles of the documents,

the ratio of its occurrence number in the document to its occurrence number in the set of documents, and

the ratio of the number of the documents including the word to the number of documents in the set of documents.